Making Publication-Quality Scientific Tables and Figures using JMP

Introduction

In this assignment you will be working with a dataset of land snail abundances and environmental data. These were collected by my research team of undergraduates in the Driftless (unglaciated) region in Wisconsin. You will be investigating this data to determine if there is a significant difference in land snail community in grassland vs forested habits. Your hunch is that that forest sites will be more species rich. We will be using this data set to practice calculating basic descriptive statistics, making tables and figures, and practicing conducting and interpreting data analyses similar to what you will be required to do with the class's data to submit as your own analysis and to present on a poster. To do this assignment you will be working with JMP, a common stats package. I have detailed instructions below but if you get stuck the JMP instructions are very easy to use. https://www.imp.com/support/help/13/Using JMP.shtml

Statistical Note: This data is roughly normally distributed and the tests we are using are relatively robust to violation of that assumption. We will have to test for normality and modify our approach if our actual research dataset has different properties.

Task 1: Open the sample data for stats practice file in JMP. Examine the data, noticing the sites, environmental data, and abundance data for each snail species – already transformed into number of individuals per m².

Descriptive statistics

Task 2: In this section you will use the column of data with the Number of Species Present grouped by Habitat and make a table. Be sure you follow the rules for making a Scientific Table that you learned in the Basic Tables and Figures assignment.

JMP instructions:

- a. Click on the column labeled "Number of Species" to highlight it.
- b. Then go to Tables>Summary.
- c. In the dialog box that opens (Figure 1) select the statistics you want to present using the drop menu that appears when you click Statistics, I suggest at least N, mean, variance, quartiles, maximum, and minimum values.
- d. In the Select Column box select the line for Habitat and then click the Group button.
- e. Click Ok to run the descriptive statistics analysis. A spreadsheet of data will then pop up with the data you requested.
- f. Take that set of data and paste it below, modify it to match the format you learned and include a Table legend.

INSERT TABLE 1 HERE:

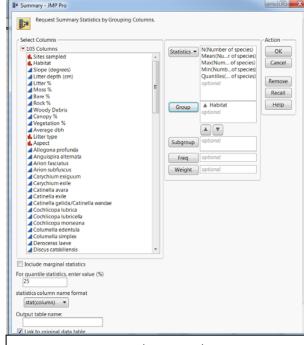


Figure 1. Descriptive (summary) statistics dialog box.

Make a Distribution graph (histogram) of Grassland and Woods species data

Task 3: Make a table with summary statistics of these datasets.

JMP instruction:

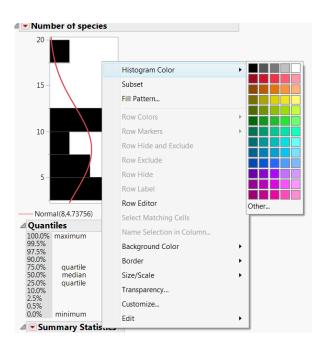
- a. To work with each subset of data you will have to make subset spreadsheets first. In your original data spreadsheet click and scroll to select (highlight) all the sites with a Grassland habitat.
- b. Click Tables>Subset and in the dialog box that pops up choose Selected Rows and click ok. That will give you a dataset of just the grassland sites to analyze.
- c. Follow that same route to also set up a subset dataset for the Woods sites.
- d. In your Grassland subset data click the distribution button . Also under Analyze> Distribution.
- e. In the Select Columns area select "Number of Species", click the "Y, columns" button to add it to the analysis. Notice, now you are only analyzing the Grassland portion of the dataset. Click OK.
- f. What appears is a histogram showing the distribution of data points in this subset of data, you also get the Summary statistics for the grassland samples.
- g. Use the layout of your Table 1 above and add columns to show grassland and woods samples separately. Include Mean, Standard Error, maximum, minimum, 75%, and 25% quartiles, and N.
- h. Paste below as Table 2. We will use this info to make graphs next.

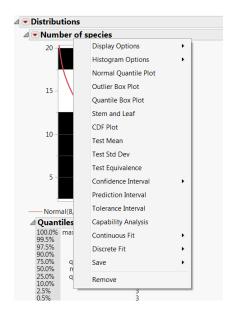
INSERT TABLE 2 HERE:

Task 4: Making Graphs of these subsets of data.

JMP instruction:

- a. In the graphical display that comes up you can also modify what types of visuals you are seeing.
- b. Click on Number of species and many display options will come up. Turn on the Normal Quantile Plot, the CDF plot, and the Stem and Leaf plots so you can see other ways of examining this data.





JMP instruction:

- a. Right click in the histogram itself to explore the options for presenting this graph.
- b. Change the bars of the histogram to solid black by using the Histogram Color tool.
- c. Go to Edit, then copy graph to get a version of the figure to paste below.
- d. Be sure to click off the extra options you explored so you just copy the histogram, paste it into this document, and add a figure legend.

PASTE DISTRIBUTION GRAPH FOR GRASSLAND SAMPLES HERE.

PASTE DISTRIBUTION GRAPH FOR WOODS SAMPLES HERE.

Make a box and whisker plot of grassland vs woods data.

Task 5. Make box and whisker plot of each subset of data, summarize results.

JMP instructions:

- a. Go back to your original dataset and click on the Graph Builder icon . Also under Graph>Graph Builder.
- b. Select the Box and Whisker Icon from the row of graph types presented (see below).



- c. Select Box Type: Quantile.
- d. From the Variables area select and drag the Habitat item and the Number of Species Present item to the drop zone.
- e. As with the distribution graph, right click on the graph area and click Edit and the Graph options to modify visual aspects of the plot.
- f. Click Edit>Copy graph and paste it below, be sure to write a complete figure legend.

PASTE BOX AND WHISKER PLOT HERE:

INTERPRETATION OF BOX AND WHISKER GRAPH: Write a 2 sentence summary of what the bar graph shows of your findings. Be sure to reference the mean values as well as the standard error.

Bar graph

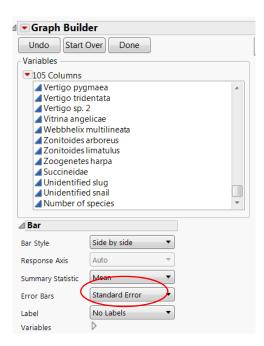
Task 6. Make bar graph of each subset of data, summarize results.

JMP instructions:

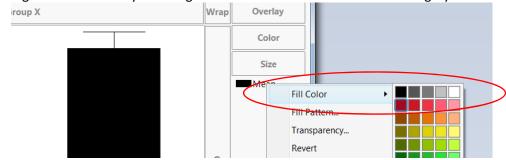
- a. Go back to your original dataset and click on the Graph Builder icon . Also under Graph>Graph Builder.
- b. Select the Bar graph Icon from the row of graph types presented (see below).



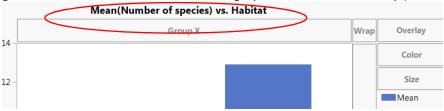
- c. From the Variables area select and drag the Habitat item to the X axis and the Number of Species Present item to the Y axis in the drop zone.
- d. As with the distribution graph, right click on the graph area and click Edit and the Graph options to modify aspects of the plot.
- e. Show standard error bars. In the Bar menu area on the left side you have options of what statistic to graph and error bars. Show error bars of standard error.



f. Right click on the Key to change the fill color of the bars to a medium gray.



g. Click on and delete the Title of the graph as this is unnecessary (and annoying) in a scientific figure.

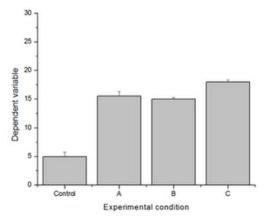


h. Click Edit>Copy graph and paste it below, be sure to write a complete figure legend.

PASTE BAR GRAPH HERE:

INTERPRETATION OF BAR GRAPH: Write a 2 sentence summary of what the bar graph indicates of your findings. Be sure to reference the mean values as well as the standard error.

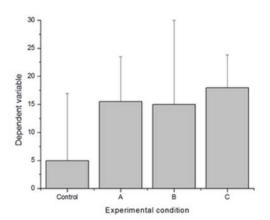
Thinking about your graphs: What do all those lines mean?



When you are graphing the average (or median or similar) values you <u>must also</u> include some measure of how much variation there is in the data and tell the reader which you are using. E.G. Standard error, Standard deviation, variance, confidence interval, etc..

Why?

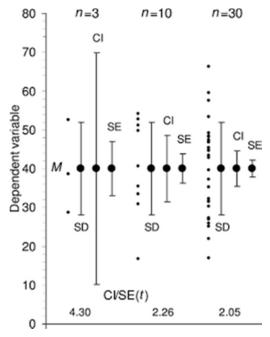
This graph makes it appear that the treatment makes a big difference compared to the control.



This graph, which plots the exact same *averages*, indicates a completely different result, depending on the measure of variation that is graphed.

The error bars could represent at least three different ways of measuring variation.

Standard deviation, standard error, or confidence interval. There is no convention for distinguishing them.

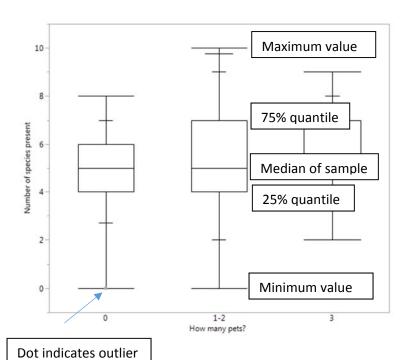


Let's compare what they could look like side by side for the same data and how they are effected by sample size. See how standard error which takes the N into the calculation shrinks with increasing sample size?

All of these are indicated on graphs the same way – with error bars. So you MUST indicate in your figure legend which one you are plotting.

In this assignment in JMP you should select the quantile button so you will be showing error bars illustrating the quantiles.

What are all those quantile lines?



Conduct a t-test on species richness and answer the questions posed

T-Test Background: How do we know if this (any) difference in results due to our treatments is significant?

We need to do a t-test. A t-test allows us to determine if two groups are significantly different. It assumes the values we are measuring are continuous and normally distributed, which our data are. A t-test works by comparing the mean values of the two treatments and taking into account the variability in the data (the variance around the means) for two groups. The formula is included for your reference, but we will use a computer program to calculate for us.

Before we run the calculation we need to look back at our hypotheses. In statistics, our baseline expectation or null hypothesis is that there are no differences between the groups, the alternative hypothesis is that we see a

difference. Rejecting the null hypothesis and thus finding support for some relationship between variables is the central activity of statistical analyses.

Variability in your data: A t-test does not just compare the means of your populations; it also includes the amount of variability (the variance).

Look at the graphs and the p-values.

Each of these illustrations show two populations with <u>average values</u> that are the same, but the variability is different. In the low variability example, the distributions for the two groups are very narrow and overlap only slightly. In the medium variability example, the distributions overlap much more, and in the high variability example the two distributions overlap almost entirely. A t-test looks at the ratio of the difference in group means to variability. Samples with higher variability with have higher p-values.

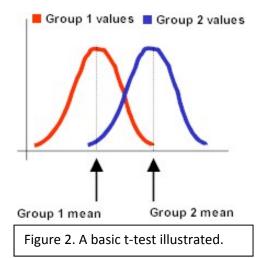
Task 7: Conduct t-test on Number of species by habitat.JMP Instructions:

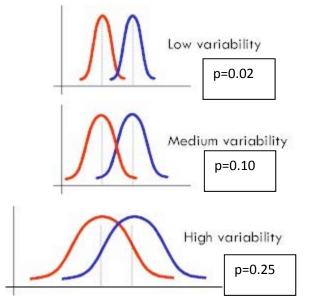
- a. Go back to your original dataset (including both grassland and woods data). Click on the Fit Y by X button, Also under Analyze>Fit Y by X. As we have a continuous Y variable but a
- categorical X variable this will do a One-way analysis. b. Add Number of species to the Y, Response area.
- c. Add Habitat to the X, Factor area. Click Ok.
- d. In the Graph that comes up, right click on the header and select "T-test"
- e. Examine the results that are calculated. The value for "Prob > |t|" is the p-value of a two-tailed test. That is what you are looking for in this example.

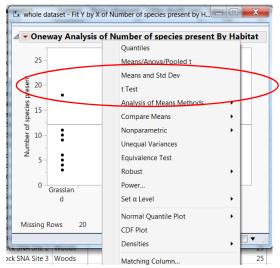
Interpreting your t-test results and p-value

We are going to gloss over the details a little here, but basically, in science,

we accept a 5% (0.05) chance that our populations are actually identical when we think they are different. In other words, we are OK with the idea that 5% of the time we incorrectly reject the null hypothesis. This significance level is







also known as the α ("alpha") for a given statistical analysis. In the statistics typically used in biology we use a p-value of 0.05. Any p-value greater than 0.05 is considered non-significant so we cannot reject our null hypothesis, a p-value lower than 0.05 is significant allowing us to reject the null hypothesis and support the alternative hypothesis.

ANSWER QUESTIONS ABOUT THE RESULT OF THIS TEST:

- 1. What is the value for $Prob > |t|^{"}$?
- 2. Is this a statistically significant result?
- 3. Does this result fail to reject or reject your null hypothesis?
- 4. Does this result support or reject your alternative hypothesis (research question)?
- 5. What does this tell you about land snail species richness in these two habitats?
- 6. What are possible alternative explanations for this results (other than you hypothesis being accurate)? Here we are trying to brainstorm. Is there anything else that could explain this pattern? This is a hugely important question whenever you are doing an experiment.

Pick an environmental variable to explore

Task 8: Select an environmental variable and determine if it significantly affects the presence of any of the snail species. Present a graph and the results of a statistical test that determine significance.

WHAT ENVIRONMENTAL VARIABLES AND SPECIES ARE YOU CONSIDERING? SUMMARIZE HERE:

PASTE GRAPH HERE:

SUMMARIZE STATISTICAL TEST AND RESULTS HERE: